

Appln No. 10/524,606
Amdt date April 23, 2008
Reply to Office action of January 28, 2008

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1-75 (Cancelled)

76. (Previously Presented) In a training data set D containing gene expression data for a plurality of genes derived from a plurality of normal cells and a plurality of diseased cells, a method for determining a plurality of emerging patterns within the gene expression data, wherein said emerging patterns can be used to test cells from a test sample for the presence of the disease, the method comprising the steps of:

determining a plurality of emerging patterns from the training data set D, wherein each of said emerging patterns comprises at least one condition based on a fixed range of said gene expression data for at least one of said plurality of genes and wherein a plurality of occurrences satisfies said at least one condition for one of said normal and diseased cells, but no occurrence satisfies said at least one condition for the other of said normal and diseased cells;

creating 2 lists, wherein: a 1st list of said 2 lists contains a frequency of occurrence, $f_1(m)$, of each emerging pattern $EP_1(m)$ from said plurality of emerging patterns that has a non-zero occurrence in data associated with said normal cells; and a 2nd list of said 2 lists contains a frequency of occurrence, $f_2(m)$, of each emerging pattern $EP_2(m)$ from said plurality of emerging patterns that has a non-zero occurrence in data associated with said diseased cells

wherein said 2 lists each provide a plurality of emerging patterns that can be used as a diagnostic tool to determine whether the cells from the test sample are normal or diseased.

77. (Previously Presented) The method of claim 76, wherein at least one of said emerging patterns comprises at least three conditions for each of said normal and diseased cells.

78. (Previously Presented) The method of claim 76, further comprising steps for:

- providing a test sample;
- testing a plurality of genes in said test sample to produce a test data set, said plurality of genes comprising at least a portion of the genes identified in said plurality of emerging patterns for said training data set D;
- determining a plurality of emerging patterns from the test data set, wherein each of said emerging patterns comprises at least one condition based on a fixed range of said gene expression data for at least one of said plurality of test genes and wherein a plurality of occurrences satisfies said at least one condition for at least one of said normal and diseased cells, but no occurrence satisfies said at least one condition for the other of said normal and diseased cells; and
- using a fixed number, k , of emerging patterns extracted from said training data, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns, calculating 2 scores, a 1st score corresponding to the data associated with the normal cells is derived from the frequencies of k emerging patterns in said 1st list that also occur in said test data; and a 2nd score corresponding to the data associated with the diseased cells is derived from the frequencies of k emerging patterns in said 2nd list that also occur in said test data; and
- using said two scores to determine whether said test cells are normal or diseased.

79. (Previously Presented) The method of claim 78, additionally comprising, prior to said determining step, a step for applying an entropy based discretization method to said training data set, to generate a cut point that defines said fixed range, such that said normal data falls within said fixed range on one side of said cut point, and said diseased data falls on an opposite side of said cut point.

80. (Previously Presented) The method of claim 79, additionally comprising applying a method of correlation based feature selection to said training data set, after said discretizing.

81. (Previously Presented) The method of claim 79, additionally comprising applying a chi-squared method to said training data set, after said discretizing.

82. (Previously Presented) The method of claim 79, wherein said emerging patterns are jumping emerging patterns.

83. (Previously Presented) A system for determining a plurality of emerging patterns within a training data set D containing gene expression data for a plurality of genes derived from a plurality of normal cells and a plurality of diseased cells, wherein said emerging patterns can be used to test cells from a test sample for the presence of the disease, the system comprising:

at least one memory, at least one processor and at least one user interface, all of which are connected to one another by at least one bus;

wherein said at least one processor is configured to:

access the gene expression data within the training data set D;

determine a plurality of emerging patterns from the training data set D, wherein each of said emerging patterns comprises at least one condition based on a fixed range of said gene expression data for at least one of said plurality of genes and wherein a plurality of occurrences satisfies said at least one condition for one of said normal and diseased cells, but no occurrence satisfies said at least one condition for the other of said normal and diseased cells;

create 2 lists, wherein:

a 1st list of said 2 lists contains a frequency of occurrence, $f_1(m)$, of each emerging pattern $EP_1(m)$ from said plurality of emerging patterns that has a non-zero occurrence in data associated with said normal cells; and

a 2nd list of said 2 lists contains a frequency of occurrence, $f_2(m)$, of each emerging pattern $EP_2(m)$ from said plurality of emerging patterns that has a non-zero occurrence in data associated with said diseased cells; and

wherein said 2 lists each provide a plurality of emerging patterns that can be used as a diagnostic tool to determine whether the cells from the test sample are normal or diseased.

84. (Previously Presented) The system of claim 83, wherein at least one of said emerging patterns comprises at least three conditions for each of said normal and diseased cells.

85. (Previously Presented) The system of claim 83, wherein said processor is further configured to:

receive data from a plurality of genes in said test sample to produce a test data set, said plurality of genes comprising at least a portion of the genes identified in said plurality of emerging patterns for said training data set D;

determine a plurality of emerging patterns from the test data set, wherein each of said emerging patterns comprises at least one condition based on a fixed range of said gene expression data for at least one of said plurality of test genes and wherein a plurality of occurrences satisfies said at least one condition for at least one of said normal and diseased cells, but no occurrence satisfies said at least one condition for the other of said normal and diseased cells; and

using a fixed number, k , of emerging patterns extracted from said training data, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns, calculate 2 scores, wherein a 1st score corresponding to the data associated with the normal cells is derived from the frequencies of k emerging patterns in said 1st list that also occur in said test data; and a 2nd score corresponding to the data associated with the diseased cells is derived from the frequencies of k emerging patterns in said 2nd list that also occur in said test data; and

determine whether said test cells are normal or diseased using said two scores.

86. (Previously Presented) The system of claim 85, wherein, prior to determining said plurality of emerging patterns, said processor applies an entropy based discretization method to said training data set, to generate a cut point that defines said fixed range, such that said normal data falls within said fixed range on one side of said cut point, and said diseased data falls on an opposite side of said cut point.

87. (Previously Presented) The system of claim 86, wherein, after applying said discretization method, said processor further applies a method of correlation based feature selection to said training data set.

88. (Previously Presented) The system of claim 86, wherein, after applying said discretization method, said processor further applies a chi-squared method to said training data set.

89. (Previously Presented) The system of claim 86, wherein said emerging patterns are jumping emerging patterns.

90. (New) In a training data set D containing mushroom characteristic data for a plurality of mushrooms derived from a plurality of edible mushrooms and a plurality of poisonous mushrooms, a method for determining a plurality of emerging patterns within the mushroom characteristic data, wherein said emerging patterns can be used to test mushrooms from a test sample to determine if said test mushrooms are edible, the method comprising the steps of:

determining a plurality of emerging patterns from the training data set D, wherein each of said emerging patterns comprises at least one condition based on a particular characteristic of said mushroom characteristic data for at least one of said plurality of mushrooms and wherein a plurality of occurrences satisfies said at least one condition for one of said edible and poisonous

Appln No. 10/524,606
Amdt date April 23, 2008
Reply to Office action of January 28, 2008

mushrooms, but no occurrence satisfies said at least one condition for the other of said edible and poisonous mushrooms;

creating 2 lists, wherein: a 1st list of said 2 lists contains a frequency of occurrence, $f_1(m)$, of each emerging pattern $EP_1(m)$ from said plurality of emerging patterns that has a non-zero occurrence in data associated with said edible mushrooms; and a 2nd list of said 2 lists contains a frequency of occurrence, $f_2(m)$, of each emerging pattern $EP_2(m)$ from said plurality of emerging patterns that has a non-zero occurrence in data associated with said poisonous mushrooms;

wherein said 2 lists each provide a plurality of emerging patterns that can be used to determine whether the mushrooms from the test sample are edible or poisonous.

91. (New) The method of claim 90, wherein at least one of said emerging patterns comprises at least three conditions for each of said edible and poisonous mushrooms.

92. (New) The method of claim 90, further comprising steps for:
providing a test sample;

testing a plurality of mushrooms in said test sample to produce a test data set, said plurality of mushrooms comprising at least a portion of the mushrooms identified in said plurality of emerging patterns for said training data set D;

determining a plurality of emerging patterns from the test data set, wherein each of said emerging patterns comprises at least one condition based on a particular characteristic of said mushroom characteristic data for at least one of said plurality of test mushrooms and wherein a plurality of occurrences satisfies said at least one condition for at least one of said edible and poisonous mushrooms, but no occurrence satisfies said at least one condition for the other of said edible and poisonous mushrooms; and

using a fixed number, k , of emerging patterns extracted from said training data, wherein k is substantially less than a total number of emerging patterns in the plurality of emerging patterns, calculating 2 scores, a 1st score corresponding to the data associated with the edible

mushrooms is derived from the frequencies of k emerging patterns in said 1st list that also occur in said test data; and a 2nd score corresponding to the data associated with the poisonous mushrooms is derived from the frequencies of k emerging patterns in said 2nd list that also occur in said test data; and
using said two scores to determine whether said test mushrooms are edible or poisonous.

93. (New) The method of claim 92, additionally comprising, prior to said determining step, a step for applying an entropy based discretization method to said training data set, to generate a cut point that defines said particular characteristic, such that said edible mushroom data falls within said fixed range on one side of said cut point, and said poisonous mushroom data falls on an opposite side of said cut point.

94. (New) The method of claim 93, additionally comprising applying a method of correlation based feature selection to said training data set, after said discretizing.
method of claim 93, additionally comprising applying a chi-squared method to said training data set, after said discretizing.

95. (New) The method of claim 93, additionally comprising applying a chi-squared method to said training data set, after said discretizing.

96. (New) The method of claim 93, wherein said emerging patterns are jumping emerging patterns.